Quantitative Analysis for the Natural Sciences
10636 Sand Point Way NE
Seattle, WA 98125
206-362-3299

MEMORANDUM

To: Gene Revelas, Integral Consulting
From: Alice Shelly and Lorraine Read
Subject: Predicted Congener Concentrations in Lower Willamette Surface Sediments
Date: December 15, 2008

This memorandum describes the process we used to generate predicted total congener concentrations from sampled total Aroclor concentrations in surface sediments in the Lower Willamette River.

The sampled surface sediment total Aroclor and total congener concentrations are right-skewed, and the variance in the regression relationship increases with concentration (see Figure 1). Natural log transformations of both variables results in a clearly linear relationship with homogeneous variance (see Figure 2). Sometimes this model is not favored because of the lack of clarity as to proper back-transformation methods and the coverage of a back-transformed confidence limit. However, we believe this model provides the best fit to these data and will provide the best predictions for total congener concentrations for sites where only Aroclors were measured.

There are two issues that preclude simple linear regression predictions for these data: measurement error in total Aroclor concentrations and the back-transformation issue. Both of these issues result in bias to the predictions from the regression. We use the Simulation-Extrapolation (*Simex*; Cook and Stefanski, 1994) method to address measurement error, and the Bradu-Mundlak correction to eliminate back-transformation bias.

<u>Naïve Model</u>

The model for the linear least squares ln-ln regression is:

$$E(C) = 0.624A^{1.04}, \qquad \text{Eq. [1]}$$

where E(C) is the expected congener concentration for a given Aroclor concentration A ($R^2 = 0.78$).

<u>Simex Model</u>

Measurement error on the independent variable in a linear regression can result in serious bias to the estimated parameters. The *Simex* method as implemented in R (Lederer and Kuchenhoff, 2006) was used to correct for the measurement error in the reported Aroclor concentrations. An estimate of measurement error in the ln-transformed Aroclors is needed to estimate the Simex model. The variance of each available pair of lab sediment splits (ln-transformed concentrations) provides an estimate of measurement error for a particular Aroclor. We used the average of these variances for each Aroclor, then summed these variances to estimate the variance of the summed Aroclors. There were two Aroclors (out of seven) for which there were no splits, so we used the average of the five Aroclor variance estimates in the sum for these two Aroclors. The square root of this sum of variances is 0.47, the standard observation error needed as input to the Simex model.

The estimated Simex model is:

$$E(C) = 0.382A^{1.15}. \qquad \text{Eq.} \qquad [2]$$

The two models (naïve and Simex) are displayed in Figure 3.

Bradu-Mundlak Correction

When back-transformed, predictions from a log-linear model are biased estimates of the mean in original units. Because of this bias, the coverage of confidence or prediction intervals is also in question. We use the minimum-variance unbiased estimators (MVUEs) originally described by Finney (1941) and developed by Bradu and Mundlak (1970; See also Cohn et al, 1989), which can be implemented as a correction to the back-transformed parameters from the regression described above. This method is available in SAS, (Powell, 1991), but we wrote a script program for use in R. The formulas for the correction follow:

$$E(C) = (0.382A^{1.15}) * g(q), \qquad \text{Eq.} \qquad [3]$$

where

$$g(q) = \sum_{j=0}^{\infty} \left(\frac{m^j (m+2j)}{\prod_{i=0}^{j} (m+2i)} \right) * \left(\frac{q^j}{j!} \right) * \left(\frac{m}{m+1} \right)^j, \quad \text{Eq.} \qquad [4]$$

$$q = \left(\frac{m+1}{2m} \right) * \left(s^2 - \hat{\sigma}^2_{\hat{C}_h} \right), \qquad \text{Eq.} \qquad [5]$$

m = the degrees of freedom from the regression (219),

s^2 = the MSE from the Simex regression (0.651),

$\hat{\sigma}^2_{\hat{C}_h} = s^2 \left[1 + \frac{1}{n} + \frac{(A_h - \bar{A})^2}{\sum_{i=1}^{n} (A_i - \bar{A})^2} \right]$ is the variance of $\hat{C}_h$, the predicted ln-congener concentration at ln-Aroclor concentration (A_h), and n = 221.

The prediction limits for an unbiased estimate of the mean in original units, (i.e., E(C), given by Eq. [3]), are then given by:

$$\left[E(C) * \exp(t_{.975,m} * \hat{\sigma}_{\hat{C}_h}),\ E(C) * \exp(t_{.975,m} * \hat{\sigma}_{\hat{C}_h}) \right] \quad \text{Eq.} \quad [6]$$

The resulting predictions and prediction intervals are displayed in Figure 4 and provided in the attached excel spreadsheet.

References

Bradu, D., and Y. Mundlak (1970). Estimation in lognormal linear models. Journal of the American Statistical Association 65:198-211.

Cohn, Timothy A., Lewis L. DeLong, Edward J. Gilroy, Robert M. Hirsch, and Deborah K. Wells (1989). Estimating constituent loads. Water Resources Research, 25(5):937-942.

Cook, J.R. and Stefanski, L.A. (1994) Simulation-Extrapolation estimation in parametric measurement error models. Journal of American Statistical Association, 89:1314-1328.

Finney, D.J. (1941). On the distribution of a variate whose logarithm is normally distributed. Journal of the Royal Statistical Society (supplement), 7:155-161.

Lederer, Wolfgang, and Helmut Küchenhoff (2006). A short Introduction to the SIMEX and MCSIMEX. R-news. Vol. 6/4, October 2006. http://www.stat.uni-muenchen.de/~helmut/r-news_WL_HK.pdf.

Powell, S. (1991) Implementation in the SAS System of the Bradu-Mundlak minimum variance unbiased estimator of the mean of a lognormal distribution. In "Proceedings of the 16th Annual SAS Users Group International Conference", pp. 1745. SAS Institute, Inc., Cary, NC.

Powell, S. (2002). Interagency Memo to Janet Spencer from Sally Powell, Senior Environmental Research Scientist, Department of Pesticide Regulation, California EPA. January 8, 2002. "Equations for predicted values and prediction limits for dislodgeable foliar residues" (*http://www.cdpr.ca.gov/docs/whs/memo/hsm02001*)

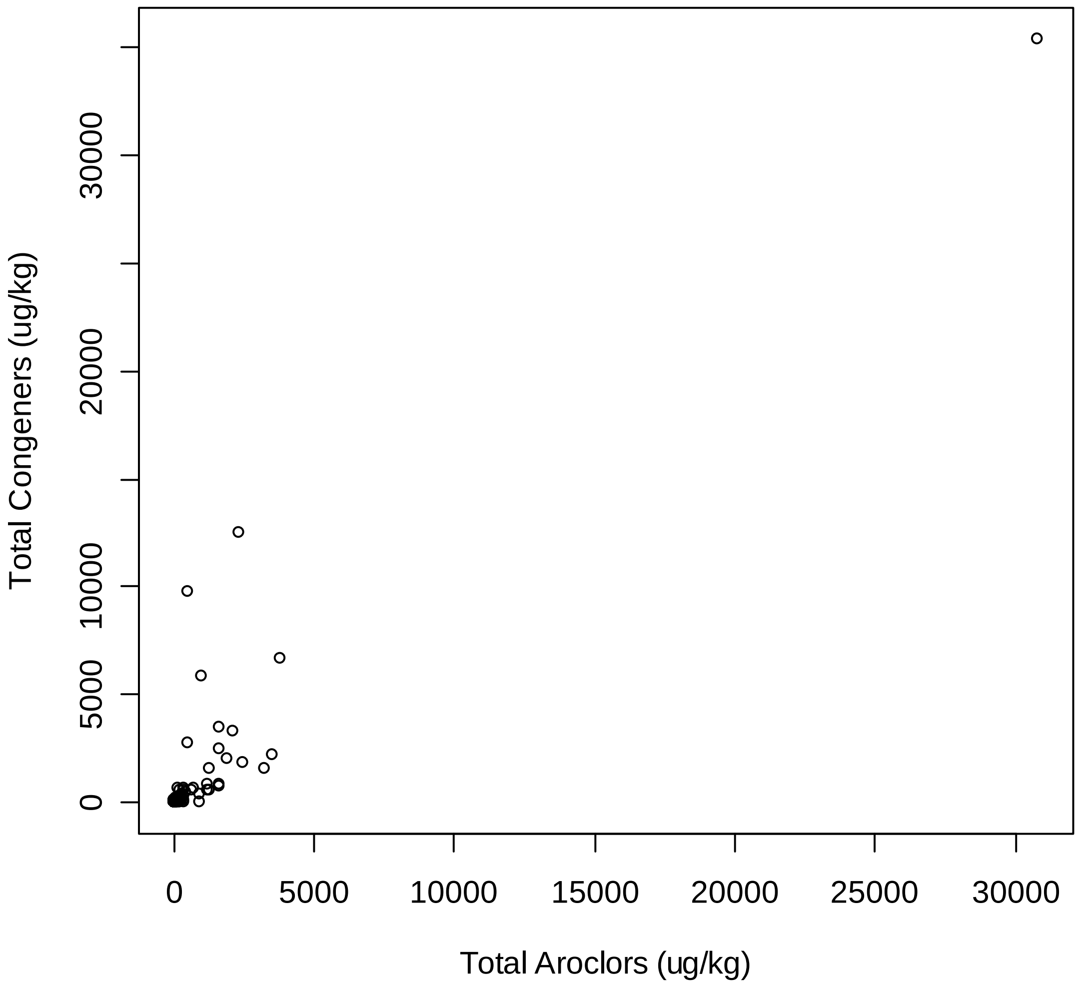

Figure 1. Measured total Aroclors vs. total congeners in surface sediments on the Lower Willamette River.

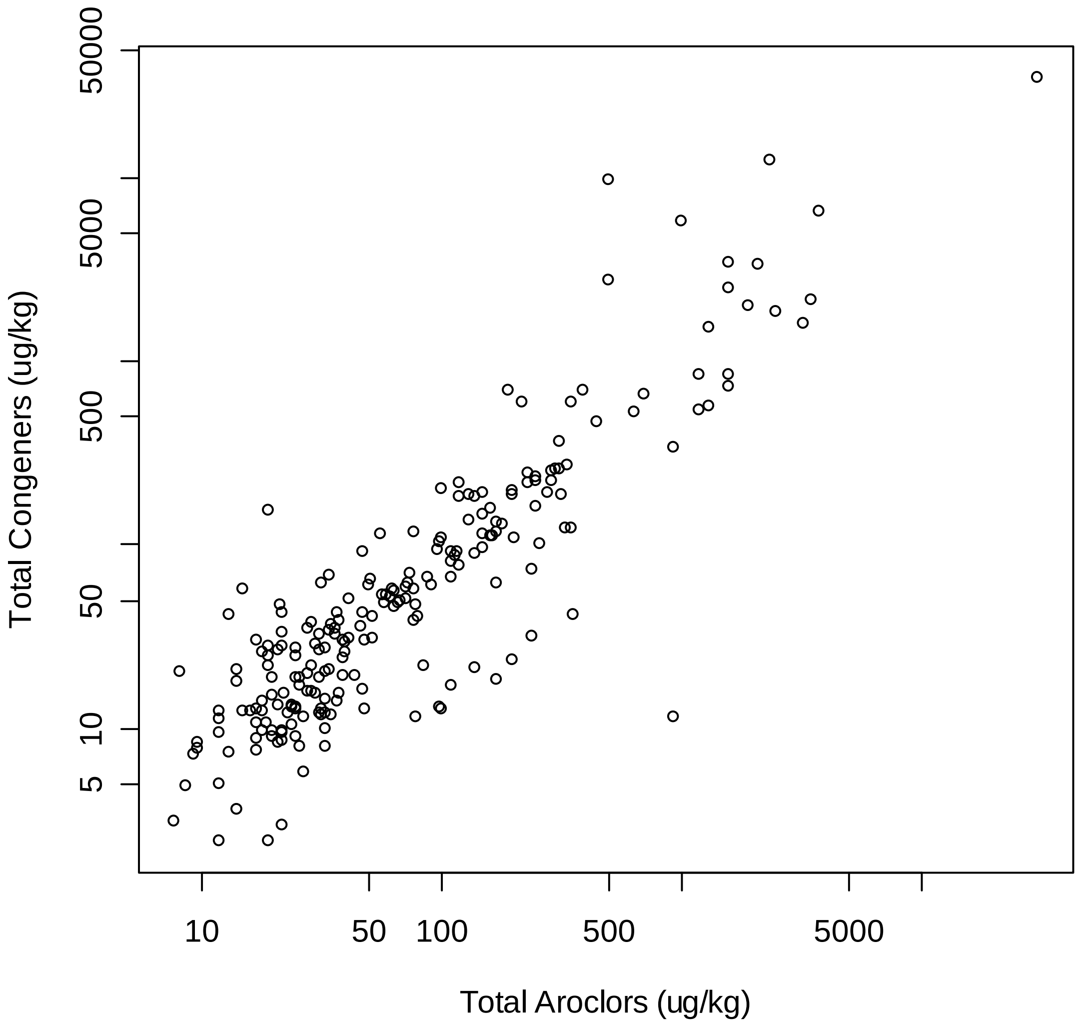

Figure 2. Total congeners as a function of total Aroclors with both variables on the log-scale.

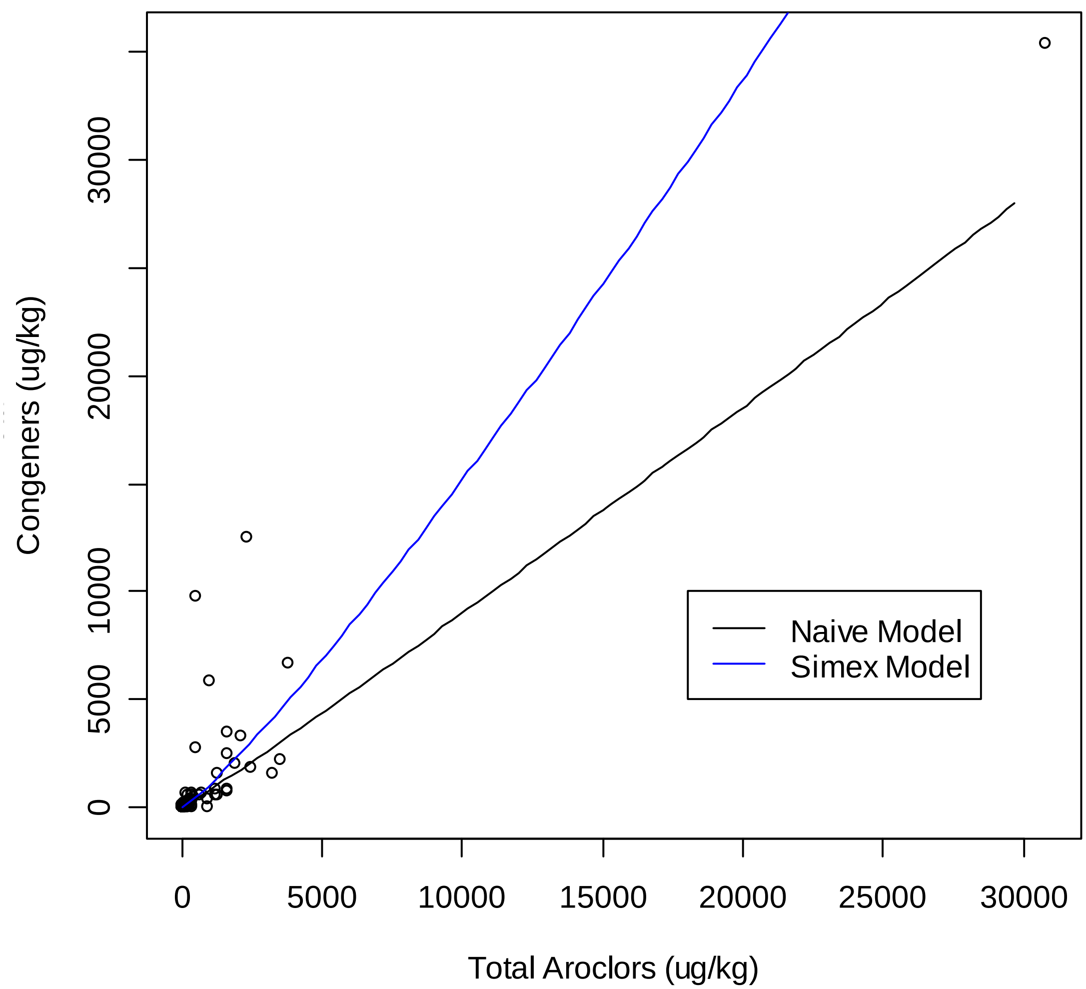

Figure 3. Naïve and Simex models compared.

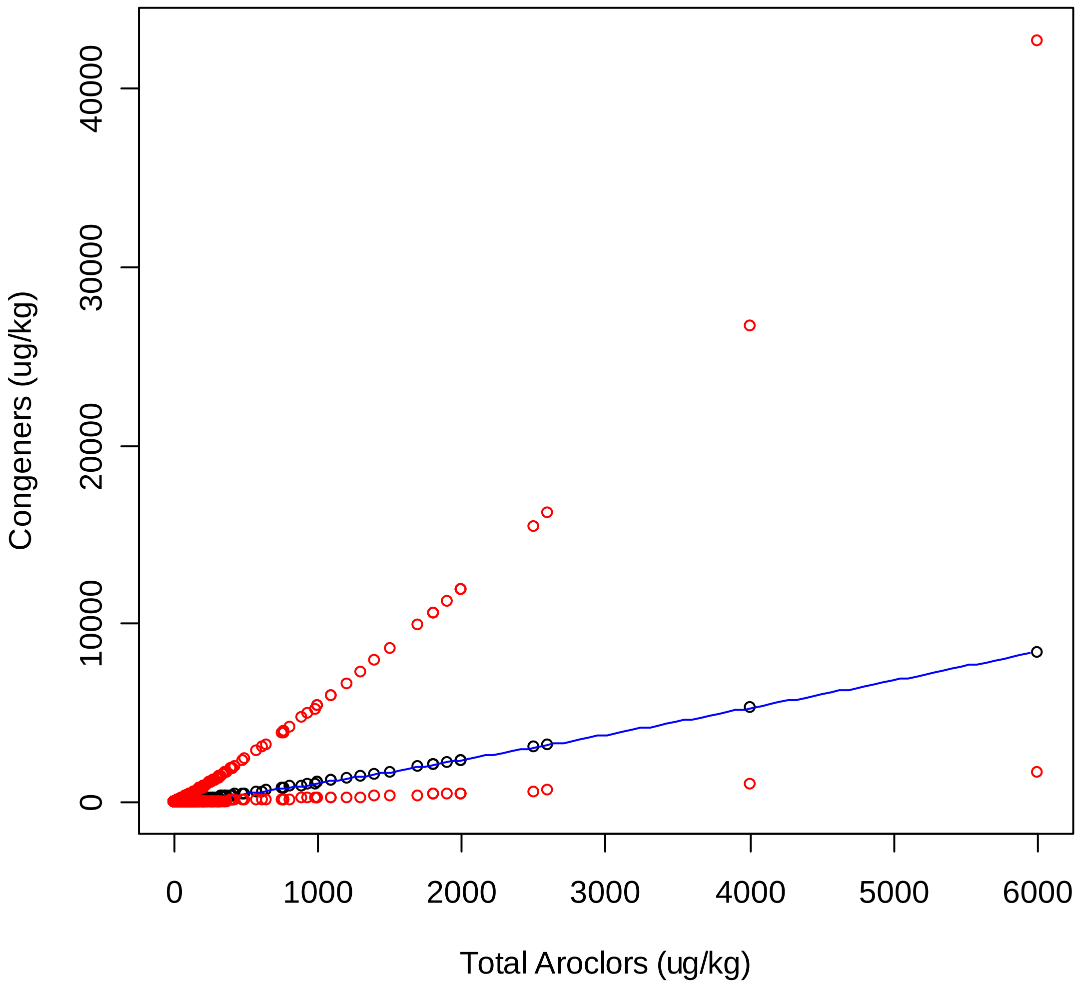

Figure 4. Bias-corrected MVUE predictions from the Simex model with 95% prediction intervals. The blue line is the Simex model fit. The black circles are the bias corrected predictions, and the orange circles are the upper and lower prediction limits.